

Mobile phone identification through the built-in magnetometers

Gianmarco Baldini*, Gary Steri*, Raimondo Giuliani*, Vladimir Kyovtorov*

*European Commission, Joint Research Centre, Ispra, Italy

gianmarco.baldini@jrc.ec.europa.eu, gary.steri@jrc.ec.europa.eu, raimondo.giuliani@jrc.ec.europa.eu,

vladimir.kyovtorov@jrc.ec.europa.eu

Abstract—Mobile phones identification through their built-in components has been demonstrated in literature for various types of sensors including the camera, microphones and accelerometers. The identification is performed by the exploitation of the small but significant differences in the electronic circuits generated during the production process. Thus, these differences become an intrinsic property of the electronic components, which can be detected and become a unique fingerprint of the component and of the mobile phone. In this paper, we investigate the identification of mobile phones through their built-in magnetometers, which has not been reported in literature yet. Magnetometers are stimulated with different waveforms using a solenoid connected to a computer's audio board. The identification is performed analyzing the digital output of the magnetometer through the use of statistical features and the Support Vector Machine (SVM) machine learning algorithm. We prove that this technique can distinguish different models and brands with very high accuracy but it can only distinguish phones of the same model with limited accuracy.

I. INTRODUCTION

The mobile phone identification based on the physical properties of its components is related to the capability to distinguish between phones of the same model but different serial numbers (intra-model identification) and between phones of different models and brands (inter-model identification). Intra-model identification is usually more difficult to achieve than inter-model identification because mobile phone manufacturers use the same materials in the same model, while different models are usually built using different materials and components.

The identification can be performed by exploiting the tiny but significant differences in the materials or the differences introduced in the manufacturing process. These differences result in small disturbances in the digital output generated by the mobile phone. They are also called *fingerprints* of the component, similarly to the fingerprints of a human being. For example, the digital camera of a mobile phone introduce in every image processed and stored in the memory of the phone, a specific feature which is called Sensor Pattern Noise (SPN). The SPN can be extracted from a sufficient number of images as demonstrated in the pioneering work by [1]. In a similar way, researchers have demonstrated the capability to identify mobile phones with different degrees of accuracy for the built-in accelerometers, radio frequency components, microphones and so on. The section II gives an overview of the research work for the different types of components and the different adopted techniques.

The identification of mobile phones through their components has various applications in security, forensics or fight against the counterfeiting of electronic products. In security, identification proofs based on physical characteristics are much more difficult to be faked and reproduced and they are intrinsically related to the component and the mobile phone itself. Fingerprints can be used to perform multi-factor authentication where physical identification is combined with cryptographic authentication (see [2]). In this case, both inter-model and intra-model identification would be applicable. Many forensics applications can exploit the knowledge of the identity of the mobile phone used in a crime scene. For example, camera identification techniques based on SPN can be applied to the images uploaded to the web by a criminal to identify the mobile phone and the criminal identity [3]. In the fight against the distribution of counterfeit products, the identification of a mobile phone through its components can be used to detect counterfeit phones, which are often built with cheaper components compared to the original ones. Cheaper components would have a different and distinct signatures from the ones used in original phones. In this case, inter-model identification would be appropriate. In fingerprinting literature, a distinction is made between *Classification* and *Verification*. In this paper, we will adopt similar definitions to what presented in [4] where Verification is the process to verify the claimed identity of a mobile phone, while Classification considers authorized mobile phone fingerprints (e.g., a reference library of known mobile phones) to create a model that best discriminates the mobile phones on the basis of the built-in magnetometers. In this paper, both processes will be implemented.

As described in the section II, there is a considerable research body for various components of the mobile phones, but there are no reported research findings for the fingerprints of the magnetometers in a mobile phone. This paper aims to address this gap.

The structure of this paper is the following: section II provides an overview of the literature of electronic device fingerprinting in mobile phones. Section III describes the methodology. In particular, the section describes the approach used to stimulate the magnetometers of the mobile phone in order to generate repeatable digital outputs. Then, the section describes the extraction of the statistical features from the digital output and the application of the Support Vector

Machine (SVM) machine learning algorithm to perform the classification and verification process. Then, section IV presents the results of the application of the classification algorithm on a set of nine phones of different models and brands including phones of the same model to test the intra-model identification. Finally, section V concludes the paper.

II. RELATED WORK

The fingerprint concept has been applied to different components of the mobile phones. There is an extensive literature on the possibility to fingerprint the digital camera of a mobile phone thanks to the imperfections present in the various components, which are part of the camera: lenses, Color Filter Array (CFA), the sensor or even the software to process the images before storing them (e.g., compression algorithms). The SPN presented in the pioneer work by [1] is the most adopted and referenced approach because of its high accuracy even for intra-model identification and its persistence in time. The SPN is based on the non-uniformity of each sensor pixel sensitivity to light. In comparison to other types of noise or imperfections having random distribution, the sensor pattern noise is a deterministic component, which stays the same for different images (in fact, it is strengthened with an increasing number of images). Even a limited set of 30-50 images can provide a SPN with good accuracy [3].

Fingerprinting of the radio frequency components of a mobile phone has also been performed by various authors for different wireless standards. Researchers have applied Radio Frequency (RF) fingerprinting to 802.11a in [2], to Global System for Mobile Communications (GSM) in [5] and to ZigBee devices in [4]. In both cases, the fingerprinting technique is based on the selection of statistical features of the collected and processed radio frequency signals emitted by the mobile phone. The statistical features are variance, standard deviation, skewness and kurtosis (e.g., [4]), which are also used in this paper. Other statistical features based on the Hilbert-Huang Transform are used in [6] for the GSM standard. RF fingerprinting usually provides very high accuracy (more than 90%) both for intra-model and inter-model classification).

The microphone of a mobile phone is another component that can be used for identification and classification. In [7], the authors use Mel-Frequency Cepstrum Coefficient (MFCC) as fingerprinting feature because it is commonly employed as a feature to characterize human speakers in audio recordings. In [7], the authors apply MFCC to the identification of the brand and model of a mobile phone on a experimental set of 14 different mobile phones. Apart from two mobile phones, which are of the same model and brand, all the other phones have different brands and models, so the analysis is mostly for inter-model rather than intra-model verification and identification. The authors used the SVM classifier for identification and verification.

Another set of features for microphone fingerprinting was used in [8], where large-size raw feature vectors are obtained by averaging the log-spectrogram of a speech recording along the time axis for each mobile phone. The authors focused

on inter-model identification and verification as they used mobile phones of various models from different brands. The features were then fed to three distinct classifiers: the Sparse representations Classifier (SRC), the SVM and Nearest Neighbour (NN). SVM provided the best performance in many configurations.

Finally, accelerometers and gyroscopes can also be used for fingerprinting as demonstrated in [9], [10] and [11]. The mobile phones are submitted to repeatable motion patterns, which stimulate the accelerometers and gyroscopes to produce specific responses. The small variations in the responses are used to create the fingerprints and distinguish between the mobile phones. Both inter-model and intra-model identification and verification were performed with very good accuracy. SVM and other classifiers were used in the cited references.

While there is an extensive literature on the identification and verification of mobile phones for different types of components, there is no reported study on the identification through magnetometers. The goal of this paper is to address this gap. We note from this survey that the use of statistical features in combination with SVM has been widely used in literature, which supports our choice to use a similar approach.

III. METHODOLOGY

A. Workflow

The goal of this section is to describe the overall methodology to perform the classification of the magnetometers built in the mobile phones, the test setup to collect the digital output (i.e., observables) of the magnetometers, the set of statistical features used to generate the fingerprints and the SVM machine learning algorithm used to perform the classification.

The overall workflow is presented in figure III-A and the single steps are described in the following subsections.

B. Stimulation of the magnetometers

The magnetometers in the phones are stimulated using a cost-effective solenoid, which is connected to the audio board of the computer. Since the goal of the experiment is to show the feasibility of mobile phone identification using low cost equipment, a consumer mass market audio board is used to stimulate the magnetometers because. A picture of the schema used to stimulate the magnetometers of the mobile phone is shown in figure 2. The audio board runs a set of three synthetic waveforms based on square waves. The waveforms are shown in figure 3. The central waveform is the reference waveform used to conduct most of the measurements; the other two are used to show the impact on performance accuracy if the density of the square waves increases or decreases. In the rest of the paper, we will identify with waveform A the first waveform (figure 3a), with waveform B the second waveform (figure 3b) and waveform C the third waveform (figure 3c). The waveforms were defined on the basis of the following considerations: a) a sharp impulse is needed to stimulate the magnetometers to generate adequate fingerprints, b) the distance between the square waves is defined on the average hysteresis values of the common mass market magnetometers and then empirically tested, c) a sequence of square waves

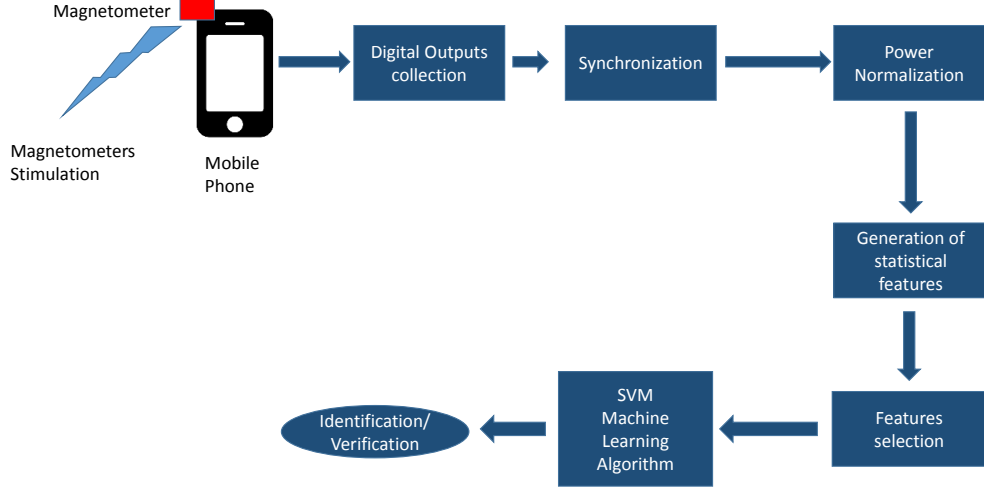


Fig. 1: Workflow for the classification of mobile phones based on magnetometers fingerprints

generates response shapes by the mobile phones, which are appropriate for the use of the statistical features (e.g., skewness and kurtosis) defined in a subsequent section of this paper. The generated audio form is then played on the audio card connected to the solenoid. Note that the sharp rise of the square wave is what generates the impulse to the magnetometer.

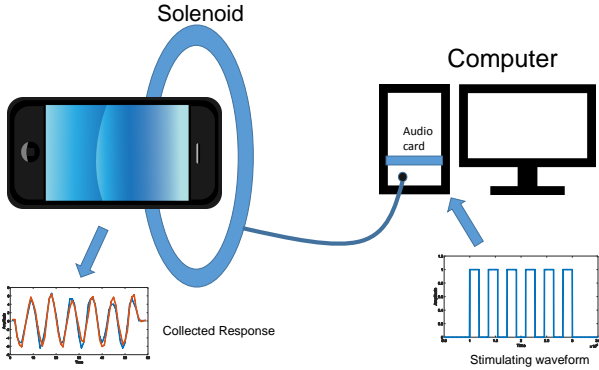


Fig. 2: Test set-up to stimulate the magnetometers of a mobile phone

Each phone is stimulated with a repetition of the 260 square waves described in 3. The value was chosen as a trade off between the need to have a statistical significant number of samples (a sample is the response of a mobile phone for a single waveform) and the time requested to stimulate the magnetometers (it is roughly one hour for the waveform B).

C. Digital outputs collection

The digital output from the magnetometers is collected using a free available application called AndroSensor that we installed on the phones. In this experiment, we use only phones supporting the android operating system but a similar study can be performed on iOS based phones. The responses from two different mobile phones to the three waveform

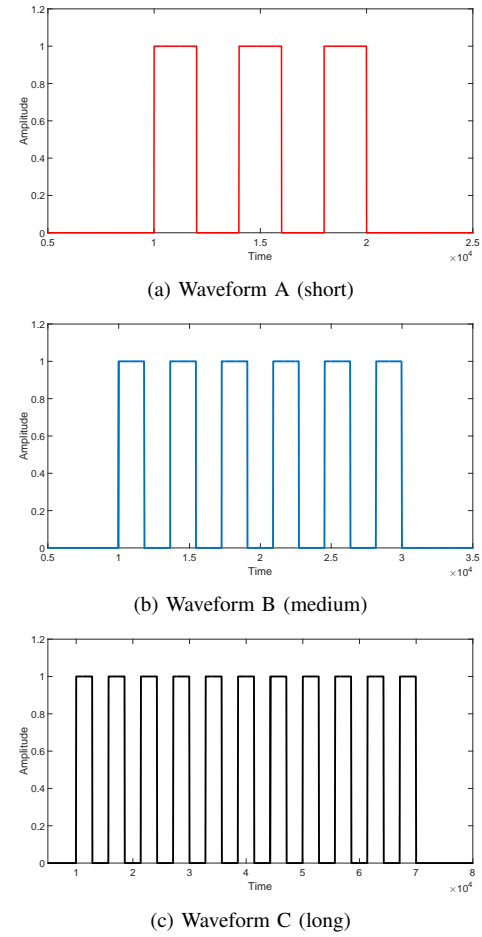


Fig. 3: Waveform patterns used to stimulate the mobile phones

after synchronization and normalization is shown in figure 4. We note that the response is quite manifold for different models of phones, which provides the unique fingerprint of

the magnetometer (and the mobile phone) as described in section IV. The collection of the data from the mobile phones was set to a frequency of 20 Hz: samples were taken by the magnetometers every 50 ms. This value was chosen because it was the lower common limit for all the mobile phones' sensors.

The list of the 9 mobile phones used in the experiments is shown in table I. Phones are from 4 different brands and 4 models. The models for which more than one phone was used are HTC One (three devices), Samsung S5 (three phones) and Sony Experia (two devices). Although the three Samsung devices are of the same model, one of them runs a different version of the Android operating system. The selection of the Samsung mobile phones with different software version was done on purpose to evaluate the impact of different software. In most cases, mobile phones were produced in different years even if they are from the same model.

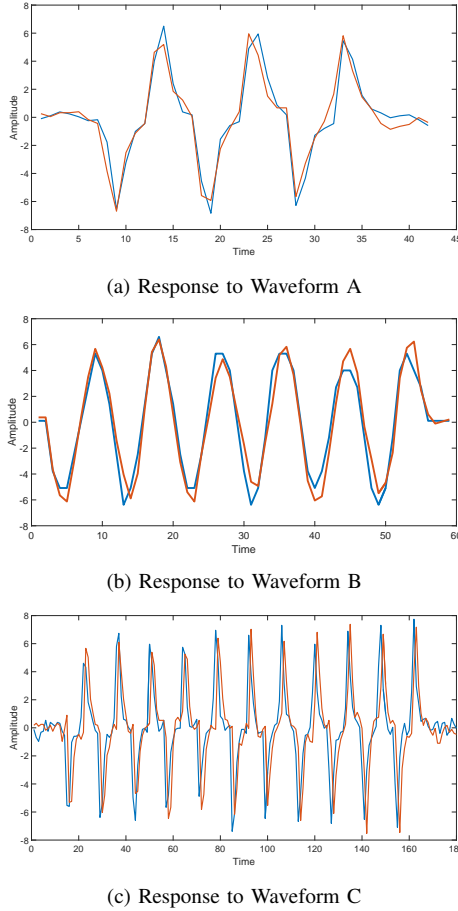


Fig. 4: Responses collected by the mobile phones for different waveforms

D. Synchronization and Normalization

The responses must be synchronized and normalized, otherwise the statistical features will generate false values, which are not dependent on the mobile phone itself but on other factors like the distance of the mobile phone from the solenoid. Synchronization and normalization is a common approach

Mobile phone model	Number of devices
HTC One	3
Huawei Ascend Mate	1
Samsung Galaxy S5 Android version 4.4	1
Samsung Galaxy S5 Android version 6.0	2
Sony Experia	2

TABLE I: List of Mobile Phones used in the experiment

in fingerprinting (see [2], [9] and [5]). The synchronization is performed using the Variance Trajectory technique. This technique is based on the calculation of the variance on a sliding window of samples, which moves along the response. The variance will increase substantially when the sliding windows meet a sharp rise or fall of the response. The rise of the variance identifies the beginning and the end of the response. This process is applied to all the 260 responses obtained in collection phase. The application of the variance trajectory was inspired by its use in RF fingerprinting to detect the start and end of the wireless communication bursts (see [2]).

The normalization is performed by applying the Root Mean Square (RMS) to each single response for each single mobile phone.

E. Statistical features

In this section, we define the statistical features used in this paper to generate the fingerprints. As described in [12], the classification of time series can be based on the representation of the time series using a set of derived properties, or features (i.e., feature-based classification). The advantage of the *feature-based* approach is that it transforms a temporal problem in a static one and the verification and identification can be therefore more computationally efficient once the statistical features have been generated. Another advantage is the presence of many statistical features, which could be used for classification. This provides a larger set of tools for identification even if there is the risk that some statistical features are similar or correlated and increasing the number of features does not provide a significant increase in accuracy. Thus a features extraction or features selection process is needed. The disadvantage of the feature based approach is that it usually requires a statistical significant number of observables from the mobile phone to perform an accurate identification and verification. This is the reason why the magnetometers must be stimulated a significant number of times (260 in our case) to support the subsequent application of the machine learning algorithm.

Since this is the first attempt in literature to fingerprint magnetometers, there is no previous research work that can recommend a set of statistical features. Then, we use the references presented in the related work section to collect features, which are used by other authors. From the fingerprint work on RF frequency components [4], [13], we selected variance, standard deviation, skewness and kurtosis; from the work on accelerometers [11], we selected entropy based features.

The definition of the features is provided in equations (1) to (6):

$$H_{ShannonEntropy} \{S_{TD}\} = - \sum_{i=1}^N (S_{TD}^2 * \ln(S_{TD}^2)) \quad (1)$$

$$H_{LogEnergy} \{S_{TD}\} = \sum_{i=1}^N \ln(S_{TD}^2) \quad (2)$$

$$\text{Standard Deviation} \{S_{TD}\} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (S_{TD} - \mu)^2} \quad (3)$$

$$\text{Variance} \{S_{TD}\} = \frac{1}{N-1} \sum_{i=1}^N (S_{TD} - \mu)^2 \quad (4)$$

$$\text{Skewness} \{S_{TD}\} = \frac{1}{\sigma^3} \sum_{i=1}^N (S_{TD}^3 - \mu^3) \quad (5)$$

$$\text{Kurtosis} \{S_{TD}\} = \frac{1}{\sigma^4} \sum_{i=1}^N (S_{TD}^4 - \mu^4), \quad (6)$$

where the term S_{TD} represent the instantaneous amplitude of the signal.

The response was also converted in the frequency domain with a Fast Fourier Transform (FFT). Then, the statistical features identified in equations (1) to (6) were applied respectively to the Phase and the Amplitude of the frequency domain representation of the response. This transformation provides a set of additional 12 features (6 for the phase of the FFT and 6 for the amplitude of the FFT) for a total of 18 features. The FFT transformation was implemented because it is a common practice applied in the surveyed papers [6] but also because the empirical evidence from the FFT has shown subtle differences, which could be exploited for classification.

In the rest of this paper and especially in the results section IV, we identify the features with the numbering of the equations: 1-6 in the time domain, 7-12 for the phase in the frequency domain and 13-18 for the amplitude in the frequency domain.

Combinations of statistical features extracted from the output of magnetometers represent their fingerprints. The goal is to select the best combination of features that can provide the optimal identification and verification accuracy. The process to achieve this goal is called features selection. Various approaches have been proposed in literature for features selection (see [12]). In this paper, we adopt a combination of a brute force approach with the Sequential Feature Selection (SFS) algorithm. This algorithm starts with a single feature or a small set of features and incrementally adds a feature at the time and measures the resulting value of metric. If the metric improves, the feature is added, otherwise another feature is checked. The process continues until the maximum (i.e., optimal) value of the metric is reached. In this paper, a metric based on the accuracy was used for the SFS algorithm. The accuracy is described in section III-F. To avoid local maximum values, a brute force approach is also performed to select one or few sets of combinations of 6 features among all the possible permutations of the features (groups of 6 on a total set of 18

features, which results in 18564 features). In the brute force approach, all the possible combinations of 6 features were calculated. The process was repeated for all the folds used in the machine learning classification (see III-F). An example for one fold is provided in the normalized graph of figure 5 where the presence of peaks (maximum values) is evident. From these optimum values, the SFS algorithm calculated which features should be added. The final result of the example of a single fold is that the best combination of features is [1 2 3 6 7 10 15 17] (where two additional features were added to one of the best sets identified from the brute force approach).

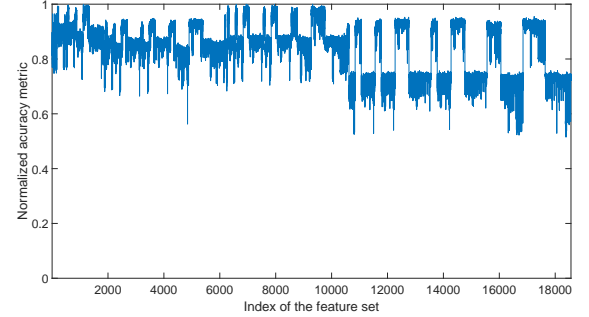


Fig. 5: Accuracy metric against the set of features for one fold

F. Machine Learning algorithms and parameters optimization

The SVM is a very well know technique in supervised machine learning and it has been used in this paper since it SVM has demonstrated its effectiveness for RF fingerprinting in the related research literature (see [14], [7] and [11]).

On the basis of a set of training samples (in the case study presented in this paper, the training samples are the features of the magnetometers responses), the SVM algorithm assigns each sample to one of two categories in the training phase. This makes SVM a non-probabilistic binary linear classifier. The resulting SVM model is a representation of the samples as points in space, mapped so that the samples of the separate categories are divided by a clear gap that is as wide as possible. In the testing phase (e.g., the verification of the identify of a mobile phone) new samples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

Here, we provide a brief formal description of the SVM algorithm to highlight the key points, which will be used in our study. Additional details on the SVM algorithm can be found in [15].

We define the n labeled examples $(x_1, y_1), \dots, (x_n, y_n)$ with labels $y_i \in \{1, -1\}$. We want to find the hyperplane H (in a proper d -dimensional space K) defined by $\langle w, x \rangle + b = 0$ (i.e. with parameters (w, b)), which satisfies the following conditions:

- 1) The scale of (w, b) is fixed so that the H plane is in canonical position,

$$\min_{i \leq n} |\langle w, x_i \rangle + b| = 1$$

- 2) The H plane with parameters (w, b) separates the +1's from the -1's, as described from the following equation,

$$y_i(< w, x_i > +b) \geq 0 \text{ for all } i \leq n$$

- 3) The H plane has a maximum margin $\lambda = 1/|w|$, i.e., minimum $|w|^2$.

We can redefine $< w, x > +b = 0$ to the following equation:

$$w \bullet \phi(x) + b = 0 \quad (7)$$

where $\phi(x)$ represents a proper mapping of x into the space K ; $w = [w_1; w_2; \dots; w_d]$ represents a d -dimensional real vector normal to H and b is a real parameter such that $|b|/||w||$ is the perpendicular distance of the origin from H . \bullet is the scalar product between two vectors.

The problem of finding the hyperplane H is an optimization problem, which can be defined on the basis of the following equation:

$$\min_{(w,b,v)} \frac{1}{2} W^T W + C \sum v_i \quad (8)$$

Where v_i are the slack variables and i is in the range of 1 to N_{Train} , which is the number of training vectors. The slack variables are subject to $v_i > 0$ and they account for the presence of classification errors. The parameter C (which we will call Box Constraint in the rest of this paper) allows the SVM user to control the weight of the classification errors in the previous equation and it is one of the two parameters to be tuned in the training process.

The second parameter to be tuned is related to the Kernel function, which is used to define the shape and format of the hyperplane. Various kernel functions are available in literature including linear, polynomial and Radial Basis Function (RBF).

In this paper, we use SVM with RBF as a kernel function because it has demonstrated its effectiveness for fingerprinting classification in [14] and other references. In addition, this kernel has a number of good features, since it can properly handle the cases in which the relation between class labels and features is nonlinear in classification problems (which is indeed our case).

The definition of the RBF is the following:

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2}, \quad (9)$$

where the γ scaling factor is the second parameter to be tuned together with C Box Constraint parameter.

To summarize, the application of SVM in this context requires the optimization of the statistical features and the parameters of the SVM and the RBF, which are the scaling factor γ from equation (10) and the Box Constraint C parameter from equation (9).

To avoid the problem of over-fitting in the training phase, where a single set of training data can contain bias not present in other data sets, a 10-fold partition is used for training and classification. In the 10-fold method, each collection of statistical fingerprints (one for each mobile phone) is divided into ten blocks. Nine blocks are used for training and one block

is *held out* for classification. The training and classification process is repeated ten times until each of the ten blocks has been *held out* and classified. In this way each block of statistical fingerprints is used once for classification and nine times for training. The final cross-validation performance statistics are calculated by averaging the results of all folds. In this way, the presence of bias in a specific training set or in portions of the training set are averaged or mitigated.

SVM is a binary classifier, and to perform classification of more than two systems as in our case (i.e., 9 mobile phones), we need to use a multi-classifier based on SVM. Two common approaches are the One Against One (OAO) and One Against All (OAA) techniques [15]. OAA involves the division of an N (i.e., 9 in our case) class data-sets into N two-class cases, while OAO involves the creation of a classification machine composed by $N(N-1)/2$ machines for each pair of systems. While OAO is more computationally intensive than OAA ($N(N-1)/2$ against N), OAA has some disadvantages especially with unbalanced training data-sets. Since we have a limited set of systems (i.e., nine) and the observables are also in limited number (i.e., 260), performance is not an issue and we decided to select OAO.

Finally, we adopted the Sequential minimal optimization (SMO) for SVM. SMO is an algorithm for solving the quadratic programming (QP) problem and it is commonly used in machine learning.

The Sequential forward selection algorithm must be based on a criterion against which the optimum value is identified. In machine learning, the following parameters are defined:

- T_p is the number of true positive matches where the machine learning algorithm has correctly identified a sample (e.g., a collected RF signal in our context) as belonging to the correct class.
- T_n is the number of true negative matches where the machine learning algorithm has correctly identified a sample as not belonging to the correct class.
- F_p is the number of false positive matches where the machine learning algorithm has identified a sample as belonging to a class while it is not true.
- F_n is the number of false negative matches where the machine learning algorithm has identified a sample as not belonging to the class while this is not true.

The combination of the different parameters can define different metrics to evaluate the effectiveness of a machine learning algorithm. In this case, we use the verification accuracy as a criterion, which is calculated as:

$$Accuracy = \frac{T_p + T_n}{TotalPopulation}, \quad (10)$$

where T_p is the number of true positives and T_n is the number of true negatives resulting from the application of the SVM machine learning algorithm to the problem of verifying that the collected fingerprints are representative of the same magnetometer evaluated in the training phase (i.e., for verification). The total population represents the total population of samples (which is the sum of T_p , T_n , F_p and F_n).

Beyond accuracy, in this paper, we will also use Receiver Operating Characteristics (ROC) and the Equal Error Rate (EER) metrics to evaluate the verification accuracy.

The ROC is calculated as ROC-like performance curve and it is generated by plotting the F_p vs. F_n as the verification threshold changes.

The EER corresponds to the point on the ROC curve where the F_p vs. F_n are equal. This metric is frequently used as a summary statistic to compare the performance of various classification systems. In general, a lower EERs indicate better system classification performance.

Finally, the confusion matrix is also used to show the results of the identification process. In the confusion matrix, each column of the matrix represents the instances of a predicted class while each row represents the instances of the actual class (and vice-versa). As in our experiments we used 9 phones, the confusion matrix shown in the results section IV has a dimension of 9*9. In the confusion matrix, the correct guesses (i.e., true positive or negative) are located in the diagonal of the table, so it's easy to inspect the table for errors, as they will be represented by values outside the diagonal. The confusion matrix is also used in this paper to define the metric employed in the SFS and for the optimization of the scaling factor and box constraint parameters. The metric is the sum of the diagonal values of the confusion matrix divided for all the values of the confusion matrix. It can be considered an extension of the accuracy metric. In general, the confusion matrix is used for classification purposes while the ROCs are used for verification.

Using the accuracy metric derived from the confusion matrix, we calculated the best set of features: [1 2 3 6 7 10 15 17] as described before and then we identified the optimum values of the scaling factor and the box constraint. This was achieved by creating a two dimensional array based on a set of values of scaling factor ranging from 2^{-8} to 2^8 and the box constraint ranging from 2^{-8} to 2^{28} . Similar ranges of values are suggested by various references like [15].

The result is shown in figure 6, which features a peak value in correspondence of a Scaling Factor equal to 2^7 (128) and a Box Constraint of 2^{22} (4194304) for a specific fold as an example. The identified set of features and optimal values of the Scaling Factor and Box Constraint for each fold are used to produce all the graphs and results in the section IV.

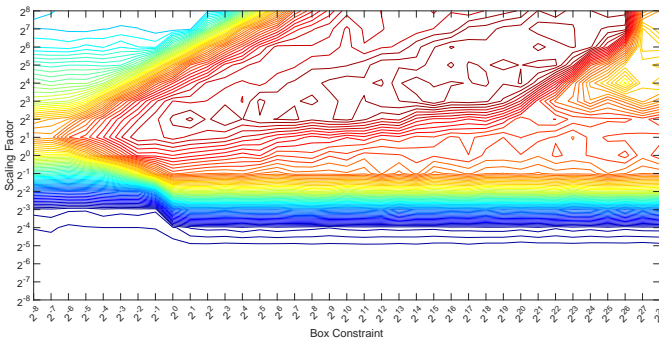


Fig. 6: Optimization of the scaling Factor and box Constraint values for one fold

	HTC One 1	HTC One 2	HTC One 3	Huawei	SAMSUNG 1 (v4)	SAMSUNG 2 (v6)	SAMSUNG 3 (v6)	SONY 1	SONY 2
HTC One 1	101	74	79	6	0	0	0	0	0
HTC One 2	36	157	33	34	0	0	0	0	0
HTC One 3	32	38	185	5	0	0	0	0	0
Huawei	0	29	0	231	0	0	0	0	0
SAMSUNG 1 (v4)	1	0	0	1	190	16	49	2	1
SAMSUNG 2 (v6)	0	0	0	0	25	197	38	0	0
SAMSUNG 3 (v6)	0	0	0	0	29	33	195	3	0
SONY 1	0	0	0	0	3	0	0	137	120
SONY 2	0	0	0	1	3	2	0	123	131

TABLE II: Confusion matrix for all the mobile phones using Support Vector Machine

IV. RESULTS

A. Results based on one day of measurements

In this section, we describe the experimental results for identification and accuracy.

Table II is the confusion matrix of all the mobile phones stimulated with the Waveform B using the SVM algorithm. From the table, we can immediately see that intra-model classification is very difficult to achieve with this approach, while inter-model classification is possible with a very good level of accuracy. More specifically, intra-brand classification is easy to achieve, because the algorithms do not confuse observables from different brands. The cross-values between HTC, Samsung and Sony are basically empty. Intra-model classification is difficult to achieve for the HTC One and SAMSUNG models, while we basically have random choice for the SONY Xperia model. We also notice that the different versions of the operating systems (version 4 against version 6) for the SAMSUNG models do not have significant impact on the classification accuracy (i.e., no effects on data collection from magnetometers, which have the same characteristics).

We also evaluated the performance of the SVM algorithm against the well known k-nearest neighbors algorithm (KNN). The resulting confusion matrix is shown in figure III, where it can be seen that the results are worst than what obtained with SVM. This can also be evaluated by calculating the ratio of the sum of the diagonal elements between the two confusion matrix (the overall sum of the elements of the matrix is the same in both cases). The result is $1545(SVM) \div 1361(KNN) = 1.135$, which shows a clear advantage for SVM (around 13% better accuracy).

The results from the confusion matrix are confirmed by the ROC curves among the different phones of different brands, models or serial number of the different models.

Figure 7 shows the ROC curves and the related EER values for verification among brands and models. ROCs are the results of a binary classification and they can be used for the verification process between a mobile phone A and B. In other words, if a mobile phone B falsely claims that it is actually mobile phone A, the algorithm should be able to verify the authenticity. ROC curves, which are more leaning against the center of the graph, indicate a lower verification accuracy. In fact, the EER parameter is the intersection of the diagonal line

	HTC One 1	HTC One 2	HTC One 3	Huawei	SAMSUNG 1 (v4)	SAMSUNG 2 (v6)	SAMSUNG 3 (v6)	SONY 1	SONY 2
HTC One 1	92	71	76	19	0	0	0	0	2
HTC One 2	47	124	29	60	0	0	0	0	0
HTC One 3	52	51	143	14	0	0	0	0	0
Huawei	4	21	1	234	0	0	0	0	0
SAMSUNG 1 (v4)	1	0	1	152	37	65	0	4	0
SAMSUNG 2 (v6)	0	0	0	0	37	174	49	0	0
SAMSUNG 3 (v6)	0	0	0	0	31	49	177	1	2
SONY 1	0	0	0	0	5	1	1	142	111
SONY 2	0	0	3	0	4	2	2	126	123

TABLE III: Confusion matrix for all the mobile phones using the k-nearest neighbors algorithm

with the ROC curve. The higher is an EER, the lower is the verification accuracy.

We notice that it is very easy to distinguish between mobile phones of different brands (HTC against Huawei, Samsung or Sony) as expected from the confusion matrix. The EER values are very low, which demonstrate a very high verification accuracy.

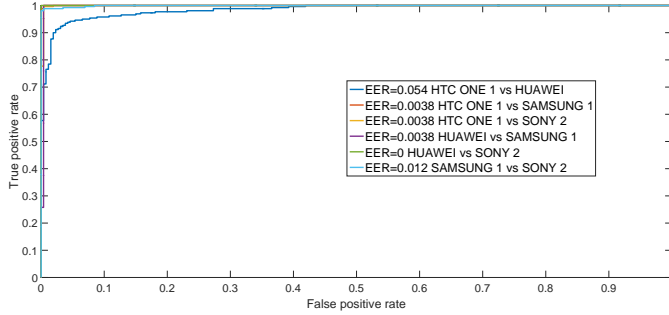


Fig. 7: Inter-model and brand verification for day one

Instead, figure 8 shows the ROC curves and the related EER values for verification between mobile phones of the same model. In our case, this set includes only the HTC One (three phones), SAMSUNG (three phones) and SONY (two phones) models. We noticed that the algorithm is able to distinguish the mobile phones only with great difficulty (EER around 0.3 and 0.2) for the HTC One and SAMSUNG phones, while we get random-choice between the SONY phones.

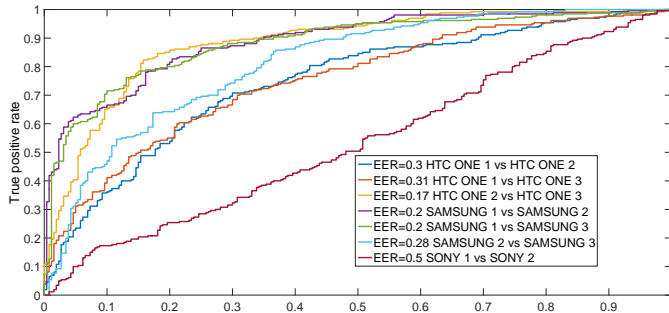


Fig. 8: Intra model verification for day one

Mobile phone model	Average EER
HTC One	0.47
Huawei Ascend Mate	0.44
Samsung Galaxy S5	0.5
Sony Experia	0.507

TABLE IV: Averaged EER among data from same phones in different days

B. Results based on different days of measurements

To ensure that the results obtained in the previous section are consistent in time, we repeated the stimulation and collection of data in different days. We collected data for other two days for all the mobile phones. Note that the measurement were not taken in consecutive days but in days separated by weeks to evaluate the stability of the algorithm for a long time-frame.

We measured the stability of the algorithm on the basis of two approaches: a) we evaluated how different are the ROCs of the verification of a mobile phone against the data sets taken in three different days for another mobile phone and b) we evaluated how similar are the data sets of the same mobile phone across the three different days. In the case a) the ROC curves and the related EER values should be quite similar, while in the case b) the EER calculated from the ROCs comparing the different data sets of the same phone should be almost random choice (EER=0.5) because they are observables taken from the same mobile phone.

The results are shown in figure 9 for the ROCs between one HTC One and the Huawei phone. We note that the ROC curves are quite near. Very similar results have been obtained for the other mobile phones.

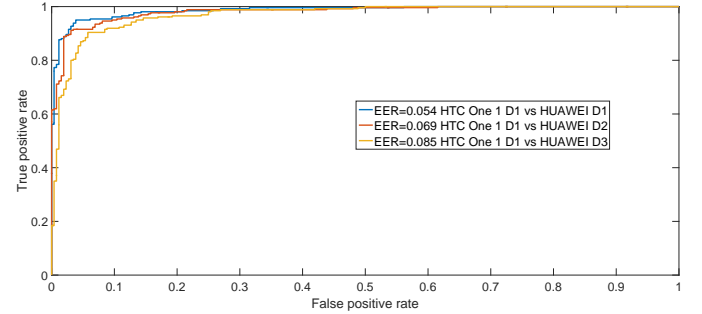


Fig. 9: ROCs of HTC One against Huawei in different days

The table IV shows the average EER for all the models used in our experiments. The three values of the EER among the combinations of days (day one against day two, day two against day three and day one against day three) have been averaged for each phone. In addition, the average among all the phones of the same model (apart from the Huawei where there is only one phone) was calculated. As expected, the resulting values of the EER are near random-choice (EER=0.5), which shows that the stability of the approach in time.

C. Results in presence of Noise

In this section, we evaluate the impact of Gaussian Noise on the verification accuracy. Additive white Gaussian noise (AWGN) was added to the digital output collected by the

magnetometers to simulate disturbances or attenuation in the propagation path between the solenoid and the mobile phone. This may be possible in a realistic application of magnetometers fingerprinting, where the position of the phone and the solenoid may not be ideal. The goal is to evaluate the loss of verification accuracy in presence of attenuation. In the simulation, AWGN with different values of Signal Noise Ratio (SNR) was added to the observables already collected. The process of adding AWGN was repeated 50 times and the results were averaged to ensure the randomness of the simulation. The results are shown in figure 10. As expected, the EER increases with the decrease of the SNR. For a value of SNR equal to 0, we obtain almost random-choice (EER=0.43).

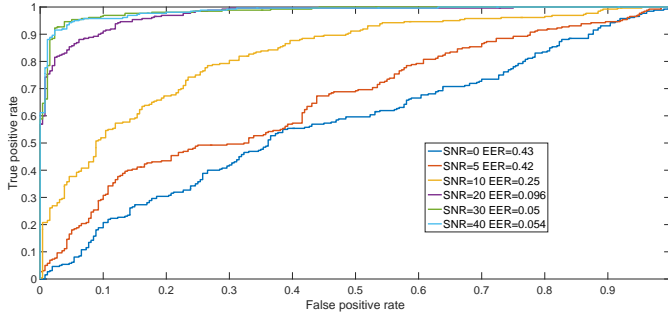


Fig. 10: Verification accuracy for different SNRs of AWGN between HTC One and Huawei

D. Results for different stimulating patterns

Finally, we evaluate the change in verification accuracy for different stimulating patterns: the waveforms from A to C described in figure 3. The goal is to evaluate if longer or more complex waveforms provide better results than shorter waveforms or vice-versa. We note that there is a trade-off because longer waveforms require longer times for the creation of the training set or for the evaluation process. In addition, the hysteresis cycle of the magnetometers require that there is a minimal time separation between the stimuli (i.e., the pulses in the waveforms of figure 3). The ROCs and the average values of EERs have been calculated for the different waveforms and the results are provided in figure 11 and 12 for the ROCs and in table V for EERs.

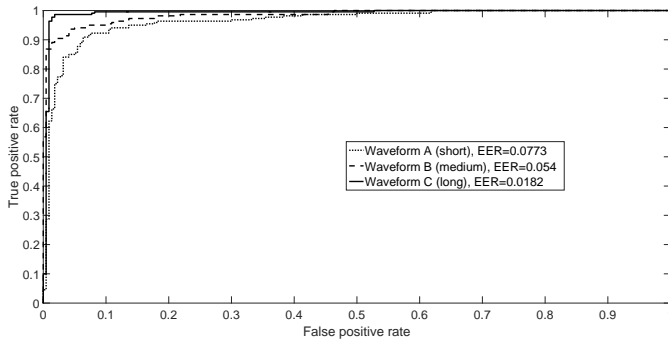


Fig. 11: ROCs between HTC One 1 and Huawei for different waveforms

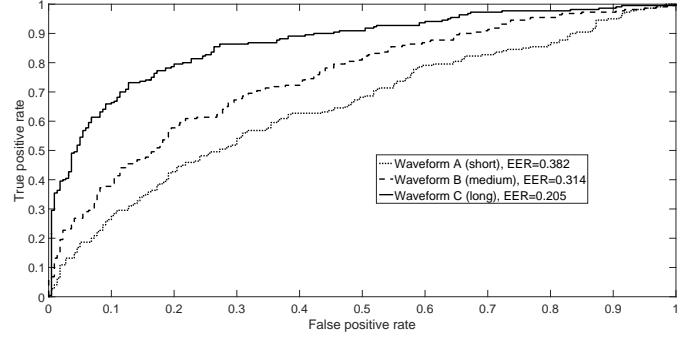


Fig. 12: ROCs between HTC One 1 and HTC One 3 for different waveforms

Intermodel	Average EER
Waveform A	0.0159
Waveform B	0.0136
Waveform C	0.00303
Intramodel	Average EER
Waveform A	0.296
Waveform B	0.334
Waveform C	0.195

TABLE V: Averaged EER for intermodel and intramodel classification

The ROCs figures (11 and 12) show the verification between the mobile phone HTC One 1 and HTC One 3 for the intra-model case and between HTC One 1 and the Huawei mobile phone for the inter-model case. We notice that the waveform C has a slightly better performance in both cases. This result is confirmed by averaging the EERs values for all the combination of mobile phones in the inter-model and intra-model cases as shown in table V where Waveform C has a clear advantage over the other two waveforms, which have similar accuracy scores.

From these results, it is clear that a longer waveform can provide a better verification accuracy at the expense of a longer training and evaluation time. Indeed, the intra-model scores for waveform C show that it would be possible to distinguish with a good level of accuracy even mobile phones of the same model and brand (i.e., intra-model).

V. CONCLUSIONS AND FUTURE DEVELOPMENTS

In this paper we have investigated the identification of mobile phones through their built-in magnetometers on a set of nine mobile phones, when they are stimulated by a solenoid with specific waveforms. We have shown that inter-model verification can be achieved with very high accuracy, while intra-model verification can be achieved with limited accuracy. We have evaluated and shown the impact of Gaussian Noise. We have also proven the stability of the verification results across different days. Regarding the application of different waveforms, the results show that accuracy can be improved with longer and more complex waveforms. Future developments will extend the analysis shown in this paper with more complex waveforms and with the application of different sets of features with the goal to obtain a better intra-model accuracy.

Disclosure Policy

The authors declare that there is no conflict of interest regarding the publication of this paper.

REFERENCES

- [1] J. Lukas, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 205–214, June 2006.
- [2] W. C. Suski II, M. A. Temple, M. J. Mendenhall, and R. F. Mills, "Using spectral fingerprints to improve wireless network security," in *IEEE GLOBECOM 2008-2008 IEEE Global Telecommunications Conference*. IEEE, 2008, pp. 1–5.
- [3] J. Fridrich, "Digital image forensics," *IEEE Signal Processing Magazine*, vol. 26, no. 2, pp. 26–37, March 2009.
- [4] T. J. Bihl, K. W. Bauer, and M. A. Temple, "Feature selection for rf fingerprinting with multiple discriminant analysis and using zigbee device emissions," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 8, pp. 1862–1874, Aug 2016.
- [5] D. R. Reising, M. A. Temple, and M. J. Mendenhall, "Improved wireless security for gmsk-based devices using rf fingerprinting," *International Journal of Electronic Security and Digital Forensics*, vol. 3, no. 1, pp. 41–59, 2010.
- [6] J. Zhang, F. Wang, O. A. Dobre, and Z. Zhong, "Specific emitter identification via hilbert-huang transform in single-hop and relaying scenarios," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 6, pp. 1192–1205, June 2016.
- [7] C. Hanilci, F. Ertas, T. Ertas, and . Eskidere, "Recognition of brand and models of cell-phones from recorded speech signals," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 625–634, April 2012.
- [8] C. L. Kotropoulos, "Source phone identification using sketches of features," *IET Biometrics*, vol. 3, no. 2, pp. 75–83, June 2014.
- [9] S. Dey, N. Roy, W. Xu, R. R. Choudhury, and S. Nelakuditi, "AccelPrint: Imperfections of Accelerometers Make Smartphones Trackable," in *2014 Network and Distributed System Security (NDSS) Symposium*, Feb 2014.
- [10] H. Bojinov, Y. Michalevsky, G. Nakibly, and D. Boneh, "Mobile device identification via sensor fingerprinting," *CoRR*, vol. abs/1408.1416, 2014. [Online]. Available: <http://arxiv.org/abs/1408.1416>
- [11] G. Baldini, G. Steri, F. Dimc, R. Giuliani, and R. Kamnik, "Experimental identification of smartphones using fingerprints of built-in micro-electro mechanical systems (mems)," *Sensors*, vol. 16, no. 6, p. 818, 2016.
- [12] B. D. Fulcher and N. S. Jones, "Highly comparative feature-based time-series classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 12, pp. 3026–3037, Dec 2014.
- [13] D. R. Reising, M. A. Temple, and J. A. Jackson, "Authorized and rogue device discrimination using dimensionally reduced rf-dna fingerprints," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 6, pp. 1180–1192, June 2015.
- [14] J. Hasse, T. Gloe, and M. Beck, "Forensic identification of gsm mobile phones," in *Proceedings of the first ACM workshop on Information hiding and multimedia security*. ACM, 2013, pp. 131–140.
- [15] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.